

# Network Energy Saving for 6G and Beyond: A Deep Reinforcement Learning Approach

Dinh-Hieu Tran, Nguyen Van huynh, Soumeya Kaada, Van Nhan Vo, Eva Lagunas, Symeon Chatzinotas

**Abstract**—Network energy saving has received great attention from operators and vendors to reduce energy consumption and CO<sub>2</sub> emissions to the environment as well as significantly reduce costs for mobile network operators. However, the design of energy-saving networks also needs to ensure that mobile users' (MUs) QoS requirements such as throughput requirements (TR). This work considers a mobile cellular network including many ground base stations (GBSs), and some GBSs are intentionally turned off due to network energy saving (NES) or crash, so the MUs located in these outage GBSs are not served in time. Based on this observation, we propose the problem of maximizing the total achievable throughput in the network by optimizing the GBSs' antenna tilt and adaptive transmission power with a given number of served MUs satisfied. Notice that, the MU is considered successfully served if its Reference Signal Received Power (RSRP) and throughput requirement are satisfied. The formulated optimization problem becomes difficult to solve with multiple binary variables and non-convex constraints along with random throughput requirements and random placement of MUs. We propose a Deep Q-learning-based algorithm to help the network learn the uncertainty and dynamics of the transmission environment. Extensive simulation results show that our proposed algorithm achieves much better performance than the benchmark schemes.

**Index Terms**—6G, Deep Q-Network, Network Energy Saving, Power Saving, Self-Organizing Networks (SONs).

## I. INTRODUCTION

Energy consumption (EC) has become a significant part of an operator's operating expenses or expenditures (OPEX). The EC on mobile networks accounts for 23% of the total operator cost [1]. Most of the power consumption comes from the radio access network (RAN) and especially from the Active Antenna Unit (AAU), with data centers and fiber optic transmission accounting for a smaller share. The power consumption (PC) of RAN can be split into two parts: the dynamic part which is only consumed when data transmission/reception (Tx/Rx) is ongoing, and the static part which is consumed all the time to maintain the necessary operation of the radio access devices, even when the data Tx/Rx is not taking place.

Therefore, network energy saving (NES) has received significant attention from vendors and operators [2], [3]. That becomes even more stringent in 6G and beyond

networks (6GBNs) with applications requiring higher speeds and lower latency, higher reliability than 5G, (e.g. extended reality/virtual reality, holographic, 3D video streaming, metaverse). Moreover, 6GBNs are becoming denser, and more antennas, bandwidth, and spectrum to adapt to mobile users' (MUs) QoS requirements. As a result, the environmental impact of 6G needs to be managed and new solutions need to be developed to improve network energy efficiency.

Recently, self-organizing networks (SONs) using Artificial intelligence/machine learning (AI/ML) have emerged as an effective solution that helps to reduce the GBSs's EC [4], [5]. With the increasing number of GBSs, mobile users (MUs), and more energy-consuming applications, more EC is inevitable. Manual management has become obsolete and inefficient in managing a large number of GBSs, and MUs for applications with heterogeneous latency, bandwidth, and rate requirements. Therefore, using an automatic operating system that uses machine learning is becoming urgent. Ericsson has shown that AI/ML can dynamically learn and then predict the MUs' traffic demand in specific areas at specific times [4]. Besides Ericsson, Nokia also prioritizes the use of AI/ML to optimize energy savings at GBSs and reduce Co<sub>2</sub> emissions [5]. Nokia shows that AI/ML can provide effective energy-saving solutions without significantly affecting the user experience. Therefore, AI/ML has become an indispensable tool for designing an automated system to operate a large number of GBSs and MUs efficiently and in an energy-efficient manner.

Recently, there have been some studies done to apply AI/ML to NES [6], [7], [8], [9]. Selcuk et al. [6] proposed a new NES algorithm called Smart Energy Efficiency using Deep Reinforcement Learning (SEEDRL) which can save up to around 9% power compared to the literature. More specifically, their algorithm aims to turn off some specific GBSs and antenna elements at each GBS to save energy without affecting the MUs's QoS. Zhou et al. [7] applied intelligent reflecting surface (IRS) as a promising technology to save energy in 6GBNs. In their study, they designed a hierarchical reinforcement learning (HRL) algorithm to control the GBSs at two levels, i.e., turn on/off the GBSs and change the transmit power of each GBSs for NES purposes.

Moreover, IRS was used in the system model to increase the average throughput of the network. In [8], Wu et al. applied transfer DRL (TDRL) to improve the NES in 6GBNs. Their purpose was to develop a novel TDRL algorithm in NES for better data efficiency compared to the baseline DRL approach. In [9], David et al. proposed a novel simulated reality of communication networks (SRCON) framework that exploited real-time network data and deployed multiple ML network models for NES. Compared to other simulators used by other operators, SRCON showed the advantages in 63% higher accuracy of the mean absolute error (MAE) and mean absolute percentage error (MAPE).

Although they give many interesting results, none of the above studies [6], [7], [8], [9] mention solutions when AI/ML makes the wrong decision to turn off one or multiple GBSs [10]. AI/ML can help us analyze the data from the past and help us to make good decision making in the future. However, with the constant change of the transmission environment in telecommunication, such as traffic demands, channel conditions, and mobility, AI/ML cannot guarantee 100% that all of its decisions are completely correct [10]. In this case, how do we cover the existing UEs and guarantee an acceptable performance before the network realizes it and turns on GBSs again. In NES, turning off GBSs to save more energy during off-peak hours is necessary but only if the MUs' QoS requirement is met. Motivated by these observations, we developed this study to ensure network robustness when wrong decision-making happens in NES for 6GBNs. Our contributions are summarized as follows:

- To the best of our knowledge, this is the first work that investigates the time domain (i.e., turn-off GBSs), power domain (i.e., GBSs' power adaptation), space domain (i.e., GBSs' antenna tilt adaptation) by applying DRL ensuring the robustness of 6GBNs when wrong decision-making to turn off GBSs happens by scheduling surrounding GBSs to support as much as possible the MUs distributed in the inactive areas. Especially, each MU requests different QoS requirements (i.e., throughput) and is located randomly around the GBS.
- We consider a multi-cell RAN network that considers the GBSs inactive/turned-off to save energy, random throughput demands from MUs, and random distribution of MUs's position. We formulate an optimization problem to maximize the achievable sum rate under the maximum allocated bandwidth, minimum and maximum GBS's transmit power, reference signal received power (RSRP) constraint, minimum MU's throughput requirement, and with a given number of successfully served MUs. The formulated problem belongs

to mixed-integer non-linear programming (MINLP) which is NP-hard.

## II. SYSTEM MODEL

We consider a downlink (DL) multi-cell system, in which the GBS is located at the center of each cell. The set of  $K$  GBSs and the set of all  $U$  MUs distributed in these cells are denoted by  $\mathcal{K} = \{1, \dots, k, \dots, K\}$  and  $\mathcal{U} = \{1, \dots, u, \dots, U\}$ , respectively. Moreover, each GBS has three different antennas to cover three sectors, the set of antenna  $i$  of each GBS is denoted by  $i \in \mathcal{I} = \{1, 2, 3\}$ .

### A. Channel Model

Let us define the distance between GBS  $k$  to MU  $u$  in 2D and 3D are respectively expressed as:

$$d_{uk}^{2D} = \|\mathbf{q}_k - \mathbf{q}_u\|, \quad (1)$$

$$d_{uk}^{3D} = \sqrt{(h_k - h_u)^2 + \|\mathbf{q}_k - \mathbf{q}_u\|^2}, \quad (2)$$

where  $h_k$  and  $h_u$  are the altitude of the GBS  $k$  and MU  $u$ , respectively. Moreover, we use the Cartesian coordinates to define the location of GBS  $k$  and MU  $u$ , respectively, i.e.,  $\mathbf{q}_j \in \mathbb{R}^{2 \times 1}$  or  $\mathbf{q} = [x_j, y_j]$  with  $j \in \{k, u\}$ .

In this work, we consider a practical system model including both light-of-sight (LoS) and non-light-of-sight (NLoS) paths that adapt to different environments, i.e., urban or rural areas. Specifically, we consider both large-scale and small-scale fading. Then, when the GBS  $k$  transmits a signal with transmit power  $P_k^i$  through antenna  $i \rightarrow$  the MU  $u$  with the effects of large-scale path loss, small-scale Rician fading, and antenna gains. Then, the received power at the MU  $u$  is given as:

$$\mathcal{P}_{uk}^i = \mathcal{P}_k^i |h_{uk}^i|^2 (d_{uk}^{3D})^{-\alpha} \mathfrak{G}_k^i(\theta, \psi) G_u, \quad (3)$$

where  $\mathcal{P}_k^i$  is the transmit power the the GBS via antenna  $i$ ,  $h_{uk}^i$  is the small-fading Rician fading,  $d_{uk}^{3D}$  is the 3D distance from GBS  $k \rightarrow$  the MU  $u$ ,  $G_u$  is the received antenna gain at the MU  $u$ , and  $\mathfrak{G}_k^i(\phi, \theta)$  denotes the transmit antenna gain of the GBS  $k$  via antenna  $i$  which is defined in the next subsection antenna model II-B.

Then, we introduce the equation to calculate the RSRP as follows [11]:

$$\tilde{\mathcal{P}}_{uk}^i = \mathcal{P}_{uk}^i - 10 \log 12 * \mathcal{N}, \quad (4)$$

where  $\tilde{\mathcal{P}}_{uk}^i$  is the Received Signal Strength Indicator (RSSI) which can be defined as the total received power at the UE including main received power, co-channel interference, and white noise,  $\mathcal{N}$  is the number of resource blocks corresponding to a channel bandwidth value.

Based on (3), the throughput (in Bits/s/Hz) at the MU  $u$  when receives the transmitted signal from GBS  $k$  is given as:

$$\mathcal{R}_{uk}^i = \log_2(1 + \Gamma_{uk}), \quad (5)$$

with:

$$\Gamma_{uk} \triangleq \frac{\mathcal{P}_k^i |h_{uk}^i|^2 (d_{uk}^{3D})^{-\alpha} \mathfrak{G}_k^i(\theta, \psi) G_u}{(\phi^{\text{RIC}} \sum_{k^* \in \mathcal{K} \setminus k} \mathcal{P}_{uk^*}^i + \sigma^2)}, \quad (6)$$

where  $\Phi^{\text{RIC}} \triangleq \phi^{\text{RIC}} \sum_{k^* \in \mathcal{K} \setminus k} \mathcal{P}_{uk^*}^i$  denotes the residual interference term after canceling all the interference from other GBS  $k^* \in \mathcal{K} \setminus k$  [12].  $\sigma^2$  is the variance of the additive white Gaussian noise (AWGN)  $n_0$  with  $n_0 \sim \mathcal{CN}(0, \sigma^2)$ .

Since  $h_{uk}^i$  is a random variable (RV), the throughput  $\mathcal{R}_{uk}^i$  is also a RV. Therefore, we aim to obtain the approximated throughput (in Bits/s/Hz) as follows [12]:

$$\bar{\mathcal{R}}_{uk}^i = \log_2 \left( 1 + \frac{e^{-E} \hat{\mathcal{P}}_{uk}^i}{\nu_{1k}^i} \right), \quad (7)$$

where  $\hat{\mathcal{P}}_{uk}^i = \mathcal{P}_k^i * (d_{uk}^{3D})^{-\alpha} * \mathfrak{G}_k^i(\theta, \psi) * G_u$ ,  $\nu_{1k}^i \triangleq (\phi^{\text{RIC}} \sum_{k^* \in \mathcal{K} \setminus k} \mathcal{P}_{uk^*}^i + \sigma^2)$ ,  $E$  is the Euler-Mascheroni constant, i.e.,  $E = 0.5772156649$  as in [13, Eq. 8.367.1].

### B. Antenna Model

In this work, we utilize a practical 3D antenna pattern defined by 3GPP [14], [15]. First, we calculate the elevation angle  $\psi$  and azimuth angle  $\theta$  between GBS  $k \rightarrow$  MU  $u$  in elevation/vertical and azimuth/horizontal planes, respectively, as follows [14], [15], [16]:

$$\theta_{uk} = \arctan \left( \frac{h_k - h_u}{d_{uk}^{2D}} \right), \quad (8)$$

$$\psi_{uk} = \arctan(\theta_u - \theta_k), \quad (9)$$

where  $\theta_u$  and  $\theta_k$  are the azimuth angle from the x-axis  $\rightarrow$  the main lobe's center of GBS  $k$  and from the x-axis  $\rightarrow$  MU  $u$ , respectively. Moreover, the  $\theta_k$  values are  $0^\circ$ ,  $120^\circ$ , and  $-120^\circ$  corresponding to three antennas in three sectors [16].

As defined in [14], [15], the azimuth and elevation gain models are respectively given as:

$$\mathfrak{G}_k^i(\psi_{uk}) = -\min \left\{ 12 * \left( \frac{\psi_{uk}}{\psi_{3\text{dB}}} \right)^2, F \right\} + \mathfrak{G}_{\text{max}}, \quad (10)$$

$$\mathfrak{G}_k^i(\theta_{uk}) = -\min \left\{ 12 * \left( \frac{\theta_{uk} - \theta_{uk}^{\text{tilt}}}{\theta_{3\text{dB}}} \right)^2, S \right\}, \quad (11)$$

where  $\mathfrak{G}_{\text{max}} = 14$  dBi is the maximum gain;  $\psi_{3\text{dB}}$  is the azimuth half-power beamwidth and equals  $70^\circ$ ;  $F$  is the front back ratio and equals 20 dB;  $\theta_{uk}^{\text{tilt}}$  denotes

the antenna down-tilt angle including both mechanical (i.e.,  $\theta_{uk}^{\text{mech}}$ ) and electrical down-tilt (i.e.,  $\theta_{uk}^{\text{elec}}$ ), when this value changes it will influence to both  $\theta_{uk}$  and  $\psi_{uk}$  values;  $\theta_{3\text{dB}}$  and  $\psi_{3\text{dB}}$  denote the azimuth and elevation half-power beamwidth and can be set to  $65^\circ$  and  $65^\circ$ , respectively;

Since Eqs. (10) and (11) represent the antenna gain along the horizontal and vertical, respectively. Therefore, to describe the antenna gain 3D pattern, we need to combine these two antenna gain patterns [14] as follows:

$$\mathfrak{G}_k^i(\theta_{uk}, \psi_{uk}) = \mathfrak{G}_k^i(\theta_{uk}) + \mathfrak{G}_k^i(\psi_{uk}). \quad (12)$$

### III. NETWORK ENERGY SAVING PROBLEM

In this section, we describe how we formulate the optimization problem. In our system, when one or more GBSs are off, the surrounding GBSs need to cover the MUs in these off GBSs. However, due to the limited resources, not all the MUs are guaranteed their QoS, e.g., the RSRP or throughput requirement.

Due to energy saving, a GBS  $k$  can be turned off, then it yields:

$$\chi_k = \begin{cases} 1, & \text{GBS } k \text{ is active,} \\ 0, & \text{GBS } k \text{ is off.} \end{cases} \quad (13a)$$

$$\chi_k = \begin{cases} 1, & \text{GBS } k \text{ is active,} \\ 0, & \text{GBS } k \text{ is off.} \end{cases} \quad (13b)$$

Inspired by a practical wireless communication system, the MUs can be served from GBS  $k$  if its RSRP is satisfied, i.e.,:

$$\vartheta_{uk} = \begin{cases} 1, & \tilde{\mathcal{P}}_{uk}^i \geq \tilde{\mathcal{P}}_{uk}^{\text{th}}, \\ 0, & \tilde{\mathcal{P}}_{uk}^i \leq \tilde{\mathcal{P}}_{uk}^{\text{th}}. \end{cases} \quad (14a)$$

$$\vartheta_{uk} = \begin{cases} 1, & \tilde{\mathcal{P}}_{uk}^i \geq \tilde{\mathcal{P}}_{uk}^{\text{th}}, \\ 0, & \tilde{\mathcal{P}}_{uk}^i \leq \tilde{\mathcal{P}}_{uk}^{\text{th}}. \end{cases} \quad (14b)$$

Then, if the MU  $u$  satisfies the RSRP condition, we add one more condition to satisfy MU  $u$ 's throughput requirements as follows:

$$\gamma_{uk} = \begin{cases} 1, & \mathcal{R}_{uk}^i \geq \mathcal{R}_{uk}^{\text{th}} \\ 0, & \mathcal{R}_{uk}^i \leq \mathcal{R}_{uk}^{\text{th}} \end{cases} \quad (15a)$$

$$\gamma_{uk} = \begin{cases} 1, & \mathcal{R}_{uk}^i \geq \mathcal{R}_{uk}^{\text{th}} \\ 0, & \mathcal{R}_{uk}^i \leq \mathcal{R}_{uk}^{\text{th}} \end{cases} \quad (15b)$$

**Definition 1.** The MU  $u$  called successfully served iff its RSRP and throughput constraints are satisfied.

Then, we introduce a new binary variable  $\pi$  such that:

$$\pi_{uk} = \begin{cases} 1, & \vartheta_{uk} = 1 \ \& \ \gamma_{uk} = 1, \\ 0, & \text{Otherwise.} \end{cases} \quad (16a)$$

$$\pi_{uk} = \begin{cases} 1, & \vartheta_{uk} = 1 \ \& \ \gamma_{uk} = 1, \\ 0, & \text{Otherwise.} \end{cases} \quad (16b)$$

Let us define  $\vartheta \triangleq \{\vartheta_{uk}, \forall u, k\}$ ,  $\gamma \triangleq \{\gamma_{uk}, k \in \mathcal{K}, u \in \mathcal{U}\}$ ,  $\pi \triangleq \{\pi_{uk}, k \in \mathcal{K}, u \in \mathcal{U}\}$ ,  $\theta \triangleq \{\theta_{uk}^{\text{tilt}}, k \in \mathcal{K}, u \in \mathcal{U}\}$ ,  $\mathcal{P} \triangleq \{\mathcal{P}_k^i, \forall k, i\}$ . Based on the above discussion, the problem of maximizing the achievable throughput at all MUs with a given number of served MUs satisfied is subject to RSRP and QoS constraints given as follows:

$$\mathcal{P} : \max_{\vartheta, \gamma, \pi, \theta, \mathcal{P}} \sum_{k \in \mathcal{K}, u \in \mathcal{U}} \log_2 \left( 1 + \frac{e^{-E} \hat{\mathcal{P}}_{uk}^i}{\nu_{1k}^i} \right) \quad (17)$$

**s.t.:**

$$\{\chi_k, \vartheta_{uk}, \gamma_{uk}, \pi_{uk}\} \in \{0, 1\}, \forall u, k, \quad (18)$$

$$\|\pi\|_1 \geq \pi_{\text{thresh}}, \forall u, k, \quad (19)$$

$$\sum_{u \in \mathcal{U}} \pi_{uk} \leq \pi_k^{\max}, \forall u, k, \quad (20)$$

$$\log_2 \left( 1 + \frac{e^{-E} \hat{\mathcal{P}}_{uk}^i}{\nu_{1k}^i} \right) \geq \mathcal{R}_{uk}^{\text{th}}, \forall u, k, \quad (21)$$

$$\mathcal{R}_{uk}^{\min} \leq \mathcal{R}_{uk}^{\text{th}} \leq \mathcal{R}_{uk}^{\max}, \forall u, k, \quad (22)$$

$$\mathcal{P}_k^{\min} \leq \mathcal{P}_k^i \leq \mathcal{P}_k^{\max}, \forall u, k, \quad (23)$$

$$d_{uk}^{\min} \leq d_{uk}^{\text{3D}} \leq d_{uk}^{\max}, \forall u, k, \quad (24)$$

$$\theta_{uk}^{\text{tilt}} \in [0^\circ, 14^\circ], \forall u, k, 8 \quad (25)$$

where (19) represents the number of served MUs that should be bigger than or equal to a threshold, i.e.,  $\pi_{\text{thresh}}$ ; Eq. (20) shows the maximum number of served MUs in one GBS is limited by a maximum value  $\pi_k^{\max}$  representing for the limited resources at that GBS; Eq. (21) denotes the rate requirement for each MU need to be higher than or equal to a threshold value, i.e.,  $\mathcal{R}_{uk}^{\text{th}}$ ; Eq. (22) shows the heterogeneous traffic demand of each MU; Eq. (23) represents the minimum and maximum transmit power of a GBS  $k$ ; Eq. (24) denotes the random locations of MUs with different distance to the GBS; Eq. (25) represents the range values of antenna tilt in one GBS.

The problem  $\mathcal{P}$  is a mix-integer nonlinear problem (MINLP), which is NP-hard. In the following section, we propose an efficient solution for applying DQN.

#### IV. DEEP Q-NETWORK-BASED SOLUTION

To overcome the uncertainty of the heterogeneous on-off GBS, MUs' traffic demand, and positions, we formulate the NES as a Markov Decision Process (MDP). More specifically, an MDP is described by a tuple  $(\mathcal{A}, \mathcal{S}, \mathcal{R})$  where  $\mathcal{A}$ ,  $\mathcal{S}$ , and  $\mathcal{R}$  are state space, action space, and immediate reward, respectively. In the following, we describe the state space, action space, and reward function. A policy  $\pi$  showing the probability of selecting an action  $a$  in state  $s$ , and can be given as  $\pi(a|s)$ .

##### A. State

It represents the configuration of the antenna tilt angle (ATA) and transmit power level of each antenna belonging to a GBS. The state can be modeled by an  $M \times 2$  matrix where  $M$  denotes the number of antennas in each GBS and 2 represents the power level and antenna tilt of each antenna.

##### B. Action Space

In each sector/cell, we have three different actions for adapting the ATA, i.e.,  $\{-1^0, 0^0, +1^0\}$ , and three different actions for changing the GBS transmit power, i.e.,  $\{-5 \text{ dB}, 0 \text{ dB}, +5 \text{ dB}\}$ . Therefore, we have a total of nine actions per sector/cell. For a GBS that includes three sectors, we have  $9^3$  actions. Then, we have an action space of  $9^C$  if we consider an area of  $C$  cells.

##### C. Immediate Reward

$\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S}$  represent the reward function, which shows the objective to optimize in (17). In this case, we consider the sum rate maximization of all MUs in the networks. Then, the reward function at state  $s$  and with action  $a$  is defined as follows:

$$\mathcal{R}_{s,a,s'} = \begin{cases} \sum_{k \in \mathcal{K}, u \in \mathcal{U}} \bar{\mathcal{R}}_{uk}^i, \vartheta_{uk} = 1 \ \& \ \gamma_{uk} = 1, & (26a) \\ 0, & \text{Otherwise.} \end{cases} \quad (26b)$$

We note that Eq. (26a) plays as a penalty in the learning process that forces our optimization problem to satisfy the RSRP and throughput constraints.

##### D. DQN-based Solution

1) *Q-learning*: Before discussing Deep Q-Networks (DQN), we would like to provide an introduction to Q-learning, which is a fundamental algorithm in RL.

First, we introduce the state value which is the total expected reward value (ERV) that we can have at a specific state:

$$V^\pi(s) = \mathbb{E} \left[ \sum_{t=0}^{\infty} \zeta^t \mathcal{R}_t | s_0 = s \right], \quad (27)$$

where  $\zeta \in (0, 1]$  denotes the discount factor in which  $\zeta$  closes to zero which means that the agent will prioritize immediate rewards and  $\zeta$  closes to one which means that the agent will prioritize the long-term reward.

Then, we discuss the value function, which is the ERV that we can have after taking a specific action  $a$  at a specific state  $s$ :

$$Q^\pi(s, a) = \mathbb{E} \left[ \sum_{t=0}^{\infty} \zeta^t \mathcal{R}_t | s_0 = s, a_0 = a \right]. \quad (28)$$

The objective of Q-learning is to maximize the expected future reward by finding the optimal policy, i.e.,  $\pi^*$ , the agent will learn to make a specific action corresponding to a specific state and iteratively converge to the optimal action-value function. As a result, the optimal  $Q^\pi(s_t, a_t)$  can be obtained by applying the Bellman equation as:

$$Q^{\pi^*}(s, a) = \mathbb{E} \left[ \mathcal{R}(s, a) + \zeta \max_{a_{t+1}} Q(s', a') | s, a \right]. \quad (29)$$

2) *Deep Q-Network*: In DQN, we use the deep neural network (DNN)  $Q(s_t, a_t; \omega)$  to obtain the optimal value of  $Q^{\pi^*}(s_t, a_t)$ , where  $\omega$  is the parameters of the DNN. The DNN inputs state  $s$  and outputs  $Q$  value for all actions. It aims to train the NN with its output to the true  $Q$  values. The DQN applies the experience relay to store the dataset from prior experience to a buffer. The prior experience is presented as tuples of  $\epsilon_t \triangleq (s_t, a_t, R_t, s_{t+1})$  into a buffer  $\mathfrak{B} \triangleq \{\epsilon_1, \dots, \epsilon_n\}$ . Then, DQN samples a random mini-batch from  $\mathfrak{B}$  to train the NN. At each iteration, the DQN agent will select an action applying  $\epsilon$ -greedy policy. After each iteration, the agent will select the best action to maximize the long-term cumulative reward.

## V. SIMULATION RESULTS

In this section, we present simulation results to evaluate the proposed joint deactivated GBSs, RSRP constraint, MU's QoS requirement, as well as the adapted antenna tilt in 6G and beyond network energy saving. We consider a system with  $U$  MUs that are randomly distributed in  $K$  GBSs following a uniform distribution. Moreover, these MUs demand a different QoS requirement, i.e., rate. Other parameters are defined as follows: number of MUs  $\|\mathcal{U}\|_1 = 80 \rightarrow 2000$  MUs, the height of the MU is  $h_u = 1.5$  meters, the height of the GBS is  $h_k = 10$  meters, mechanical down-tilt value is  $\theta_{uk}^{\text{tilt}} = 0^\circ$ , electrical down-tilt value is  $\theta_{uk}^{\text{tilt}} = [0^\circ, 14^\circ]$ , maximum gain of the antenna  $\mathfrak{G}_{\text{max}} = 20$  dBi, minimum and maximum transmit power of the GBS are  $P_k^{\text{min}} = 0$  dBm and  $P_k^{\text{max}} = 45$  dB, azimuth and elevation half-power beamwidth and can be set to  $65^\circ$  and  $65^\circ$ , respectively, the maximum queue of the buffer in memory relay in DQN is 20000. The first benchmark scheme is maximum-action-based algorithm (a.k.a. Max in the below figures), where the agent always chooses to increase the highest antenna tilt and transmit power of GBS for each action, i.e.,  $+1^\circ \in \{-1^\circ, 0^\circ, +1^\circ\}$ ,  $+5 \text{ dB} \in \{-5 \text{ dB}, 0 \text{ dB}, +5 \text{ dB}\}$ . The second benchmark scheme is a random-based algorithm (a.k.a. Random in the below figures), where the agent always chooses random values of antenna tilt and transmit power of GBS for each action, i.e.,  $-1^\circ|0^\circ|+1^\circ \in \{-1^\circ, 0^\circ, +1^\circ\}$ ,  $-5 \text{ dB}|0 \text{ dB}|+5 \text{ dB} \in \{-5 \text{ dB}, 0 \text{ dB}, +5 \text{ dB}\}$ .

In Fig. 1, we show the convergence average reward of our proposed DQN-based achievable throughput maximization problem with a given number of served MUs when the learning rate is 0.001, the batch size is set to 32 data points, and with 20000 iterations. It can be observed from Fig. 1 that as the increasing of a number of iterations of training, the agent improves the achievable throughput by increasing the reward values. In early iterations, i.e., the number of interactions is less than 10000, we see a large fluctuation in the average

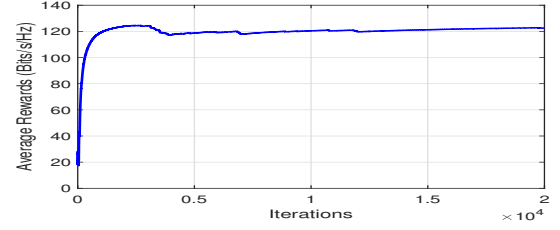


Fig. 1: Average rewards vs. number of iterations.

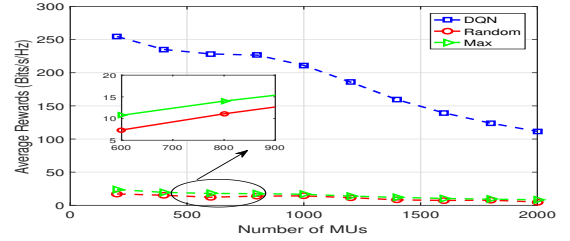


Fig. 2: Average reward per MU vs. number of MUs.

reward, it is because the agent is still spending more effort on the exploration phase to obtain the best policy. After that, the average reward increases gradually because the more exploration is applied, the better policy is obtained, which reduces the fluctuation. Another observation is that our proposed model converges to a saturated value after around 20000 iterations when the policy is quite stable. Therefore, it shows the correctness of our proposed DRL-based approach.

To investigate the influence of different numbers of MUs values on the reward  $\mathcal{R}_{s,a,s'}$ , Fig. 2 illustrates the average rate of each MU over different MUs numbers, ranging from 200 to 2000. First, it is easy to see that our proposed DQN-based solution achieves a much better sum throughput compared to the Random and Max schemes. This shows the advantages of our proposed DQN-based solution. Second, we observe that the percentage gap between the DQN-based solution to the Random and Max becomes larger with the increasing number of MUs values. For example, when  $\|\mathcal{U}\|_1$  equals 200, the average throughput of the DQN, Random, and Max are 254.67 Bits/s/Hz, 17 Bits/s/Hz, and 23.87 Bits/s/Hz, respectively. It means that the DQN solution performs better at 93.32% and 90.627% as compared to the Random and Max schemes, respectively. When  $\|\mathcal{U}\|_1$  equals 2000, the average throughput of the DQN, Random, and Max are 111.52 Bits/s/Hz, 4.925 Bits/s/Hz, and 8.36 Bits/s/Hz, respectively. It means that the DQN solution performs better at 95.58% and 92.5% as compared to Random and Max schemes, respectively. It can be explained by the fact that when we select

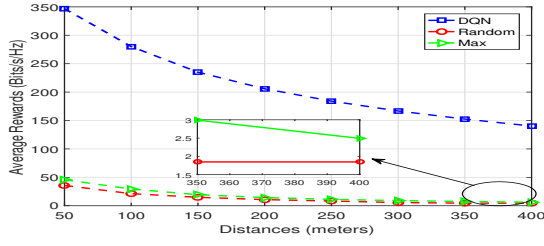


Fig. 3: Average reward vs.  $d_{uk}^{3D}$ .

random/maximum values of antenna tilt and transmit power at the GBS, it creates more interference to other MUs. i.e.,  $\phi^{\text{RIC}} \sum_{k^* \in \mathcal{K} \setminus k} \mathcal{P}_{uk^*}^i$  in Eq. (6). Therefore, it will significantly reduce the obtained throughput at each MU and reduce the sum throughput at the objective function in (17).

In order to show the impact of the average distance from GBS  $\rightarrow$  MUs (meters), Fig. 3 shows the sum throughput of all MUs over different average distances from GBS  $\rightarrow$  MUs (meters) ranging from 20 meters to 400 meters. More specifically, the distance value ranges from  $d^{\min} \rightarrow d^{\max}$ . For simplicity, we only present the  $d^{\max}$  value in Fig. 3. For example, when the  $d_{uk}^{3D}$  equals 50, the distance from GBS  $\rightarrow$  MUs ranges from 20 meters to 50 meters. Moreover, when the  $d_{uk}^{3D}$  equals 100, the distance from GBS  $\rightarrow$  MUs ranges from 50 meters to 100 meters and so on. First, we see that the sum throughput of all MUs is exponentially decreasing with a higher distance from GBS  $\rightarrow$  MUs (meters). It is because the distance from GBS  $\rightarrow$  MUs (meters)  $d_{uk}^{3D}$  is placed in the denominator in the Eq. (6). Therefore, when the  $d_{uk}^{3D}$  increases, it will significantly decrease the received throughput  $\mathcal{R}_{uk}^i$  of each MU  $u$ , and then influence the sum throughput of all MUs in the networks. Second, we observe that our proposed DQN-based method still outperforms Random and Max schemes. The difference in average reward performance decreases with increasing distance from 30 meters to 200 meters. More specifically, when  $d_{uk}^{3D}$  ranges from 20  $\rightarrow$  50 meters, the DQN performs better at 89.71% and 86.78% compared to the Random and Max schemes, respectively. It is because the higher distance value significantly reduces the achievable throughput  $\mathcal{R}_{uk}^i$  in Eq. (6) while the available resource is limited, i.e., bandwidth  $\mathcal{B}_{uk}$ .

## VI. CONCLUSIONS

In this paper, we have studied the mobile cellular networks in 6GBNs where one or more GBSs are inactive for energy-saving purposes. Moreover, we developed a novel Deep Q-learning-based algorithm that jointly optimizes the antenna tilt, power adaptation, MUs' RSRP, and

MUs' throughput requirements to maximize the sum rate with a given number of served MUs satisfied. The proposed method not only solves the problem efficiently under dynamic and uncertain environments but also deals with the high dimensional state space and action spaces. The extensive simulation results showed that our proposed algorithm provides a big gain with different numbers of MUs and distance between GBS  $\rightarrow$  MU values.

## REFERENCES

- [1] M. Oikonomakou et al., "A Power Consumption Model and Energy Saving Techniques for 5G-Advanced Base Stations," 2023 IEEE ICC Workshops, Rome, Italy, 2023, pp. 605-6103.
- [2] Ericsson, "A technical look at 5G energy consumption and performance," <https://www.ericsson.com/en/blog/2019/2/breakingthe-energy-curve-5g-energy-efficiency>, 2019.
- [3] Huawei, "5G Power: Creating a Green Grid That Slashes Costs, Emissions and Energy Use," 2020.
- [4] Ericsson, "Why AI-powered RAN is an energy efficiency breakthrough," <https://www.ericsson.com/en/blog/2023/1/ai-powered-ran-energy-efficiency>, Jan. 2023.
- [5] Nokia, "STC uses Nokia SON to improve RAN energy efficiency with AI powered energy savings management," <https://www.nokia.com/customersuccess/stc-uses-nokia-son-to-improve-ranenergy-efficiency-with-ai-powered-energy-savingsmanagement/>, Nov. 2023.
- [6] Selcuk Bassooy et al., "SEEDRL: Smart Energy Efficiency Using Deep Reinforcement Learning for 6G Networks," In 2023 IEEE Globecom Workshops (GC Wkshps), pages 732-737. IEEE, 2023.
- [7] Hao Zhou et al., "Hierarchical reinforcement learning for RIS-assisted energy-efficient RAN," In GLOBECOM 2022-2022 IEEE Global Communications Conference, pages 3326-3331. IEEE, 2022.
- [8] Di Wu et al., "Energy Saving in Cellular Wireless Networks via Transfer Deep Reinforcement Learning," In GLOBECOM 2023-2023 IEEE Global Communications Conference, pages 7019-7024. IEEE, 2023.
- [9] David López-Pérez et al., "Data-driven Energy Efficiency Modelling in Large-scale Networks: An Expert Knowledge and ML-based Approach," IEEE Transactions on Machine Learning in Communications and Networking, 2024.
- [10] Alexander S Rich and Todd M Gureckis, "Lessons for artificial intelligence from the study of natural stupidity," Nature Machine Intelligence, 1(4):174-180, 2019.
- [11] rfwireless world. LTE RSRP Calculator | LTE RSRQ Calculator | RSSI, RSRP, RSRQ. <https://www.rfwireless-world.com/calculators/LTERSRP-and-RSRQ-calculator.html>, 2024.
- [12] Dinh-Hieu Tran et al., "UAV relayassisted emergency communications in IoT networks: Resource allocation and trajectory optimization," IEEE Transactions on Wireless Communications, 21(3):1621-1637, 2021.
- [13] Izrail Gradshteyn et al., "Table of integrals, series, and products," Academic press, 2014.
- [14] Ericsson, "R1-083070: Refined Antenna Model for LTE Advanced Evaluations," [https://www.3gpp.org/ftp/tsg\\_ran/WG1\\_RL1/TSGR1\\_54/Docs](https://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_54/Docs), 2008.
- [15] 3GPP, "3GPP TR 36.814: Further advancements for E-UTRA physical layer aspects (Release 9)," <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2493>, 2017.
- [16] Ingo Viering et al., "A mathematical perspective of self-optimizing wireless networks," In 2009 IEEE International Conference on Communications, pages 1-6. IEEE, 2009.